

Г.И. Кустова

СЕМАНТИЧЕСКАЯ РАЗМЕТКА В ЭЛЕКТРОННЫХ КОРПУСАХ И ЭЛЕКТРОННЫХ СЛОВАРЯХ*

Один из видов разметки в Национальном корпусе русского языка (далее – НКРЯ или Корпус) – семантическая. Семантическая разметка дает возможность пользователю составлять содержательные запросы при поиске примеров употребления слова в определенном значении, а разработчикам Корпуса – создавать семантические фильтры для автоматического снятия неоднозначности¹. При составлении запросов и фильтров используются семантические ограничения на заполнение валентностей предикатных слов. Так, примеры употребления глагола *светить* в прямом значении можно получить по запросам: (1) *светить* + S | g:concr & t:tool:device:light (‘инструменты и приспособления; свет’); выдача: *светит коптилка, фонарик, прожектор, лампа; светил зажигалкой, фонариком* и т.п.; (2): *светить* + S | g:concr & t:space (‘пространственные объекты’); выдача: *светит солнце, луна* и т.п. Если же пользователь хочет получить примеры с переносным значением *светить* (ср. *Ему светит срок*), он должен указать в запросе другие семантические признаки актанга, а еще лучше – всю конструкцию, в которой это значение реализуется: *dat + светить* + NUM; выдача: *После окончания вуза мне «светит» 1 год срочной службы; За фальсификацию итогов голосования нерадивым или нечистоплотным работникам «светит» три года.*

Однако существительные в текстах выступают не только в функции валентных распространителей предиката, но и в статусе «свободных синтаксем» (термин Г.А. Золотовой²), наречных (не

* Работа выполняется при поддержке РГНФ, проект № 11-04-00223а и № 11-04-00488а. Литературные примеры взяты из Национального корпуса русского языка. Из соображений экономии места примеры приводятся в сокращении и без ссылки на источник.

¹ Кустова Г.И., Толдова С.Ю. Национальный корпус русского языка: семантические фильтры для разрешения многозначности глаголов // Труды междунар. конф. «Корпусная лингвистика–2008», СПб, 2008.

² Золотова Г.А. Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса. М., 1988.

обусловленных глагольным управлением) падежей (по Е. Куриловичу³) – т.е. невалентных распространителей отдельных слов или всего предложения в обстоятельственных и определительных конструкциях. Наибольший интерес представляет языковое поведение непредметных существительных, обозначающих ситуации, признаки, параметры. Предложные конструкции с такими существительными – далее будем называть их обороты – могут выступать как самостоятельные предикаты, сходные с полноценными предикатами – глаголами – по семантике и по поведению.

Предложные обороты с абстрактными существительными, по существу, являются особыми лексическими единицами⁴. Во-первых, такие существительные образуют обороты отнюдь не с любыми предлогами (речь идет о непроизводных предлогах). Многочисленны обороты с предлогами *в* (*в плену, в растерянности*), *на* (*на деле, на дежурстве; на всякий случай*), *под* (*под арестом, под охраной*), *при* (*при встрече, при свете, при жизни*), *с* (*с успехом, с недоверием; со скуки, с перепугу*), *без* (*без причины, без разбору*), *из* (*из страха, из мести*), *по* (*по ошибке, по совету*), *для* (*для смеха*), *за* (*за заслуги; за ненадобностью*); а предлог *у*, характерный для оборотов с личными существительными (*У соседей праздник; У меня новость*) и для локативных конструкций (*У забора – скамейка*), с абстрактными существительными оборотов практически не образует. Во-вторых, значения таких оборотов некомпозициональны (идиоматичны), а сами обороты фразеологичны: список употребимых в них лексем, конечно, не является жестко фиксированным, но не является и свободным. Например, исходя из существования оборота *для храбрости* (*выпил для храбрости*), можно предположить, что обнаружится и

³ Курилович Е. Проблема классификации падежей // В его кн.: Очерки по лингвистике. М., 1962.

⁴ Иорданская Л.Н., Мельчук И.А. Смысл и сочетаемость в словаре. М.: Языки славянских культур, 2007; Иомдин Л.Л. Многозначные синтаксические фраземы: между лексикой и синтаксисом // Компьютерная лингвистика и интеллектуальные технологии. Труды междунар. конф. «Диалог-2006». М., 2006; Кустова Г.И. Обстоятельственные группы типа *во всяком случае* в современном русском языке // Инструментарий русистики: корпусные подходы. Slavica Helsingiensia 34. Хельсинки, 2008.

оборот для смелости. И действительно, в Корпусе находим соответствующие примеры: *Для смелости сдвинул кепчонку вбок; Для смелости, чтобы выглядеть более раскованным, в самолете пригубил несколько грамм...; Маленков вскочил со стула, кашлянул для смелости* (всего – 17 вхождений). Однако на сочетание для *отваги* примеров в Корпусе нет, хотя это существительное – из той же семантической группы. Таким образом, и факт сочетаемости с тем или иным предлогом (т.е. списки оборотов), и значения оборотов должны фиксироваться в словаре. Поскольку ни обычные «номинативные», ни обычные фразеологические словари не рассчитаны на последовательную фиксацию подобных оборотов, их целесообразно описывать в специальном неноминативном словаре⁵. В нем обороты образуют уже новые классы, не (полностью) совпадающие с семантическими классами существительных, и получают специальную разметку, позволяющую пользователю осуществлять поиск по новым классам.

Чтобы понять, какой материал целесообразно помещать в такой словарь и какую информацию сможет получить пользователь, рассмотрим особенности поведения предложных конструкций на примере оборотов с предлогами *В* и *НА*. Эти обороты являются наиболее многочисленными по лексическому составу – в них участвует несколько сотен существительных (ср.: *в разведке, в карауле, на вахте VS. на танцах, на рыбалке VS. в плену, в заключении VS. в угоне, в розыске VS. в услужении (у кого), на содержании (у кого) VS. в бешенстве, в замешательстве, в восторге VS. в неведении VS. в дремоте, в забытии, в бреду VS. в тесноте, в толчее, в духоте VS. в / на холоде / на ветру VS. в нужде, в бедности, в роскоши* и др.) – и наиболее разнообразными по синтаксическим функциям.

Внутри этого класса выделяются подклассы, из которых самыми крупными (порядка ста единиц в каждом подклассе) и наиболее интересными для анализа являются СИТУАЦИИ и ЭМОЦИИ (в парадигму эмоций входят и другие предлоги).

⁵ Кустова Г.И. О «неноминативных» электронных словарях // Компьютерная лингвистика и интеллектуальные технологии. Труды международной конференции «Диалог-2008». М., 2008; Словарь наречий и служебных слов русского языка. Сост. Бурцева В.В. М., 2005.

Ситуации. С семантической точки зрения обороты типа *на параде, в отъезде, на дежурстве, в карауле* и т.д. обозначают ситуации (*Полк на параде; Начальник в отъезде* и т.п.). Существительные, участвующие в оборотах-ситуациях, относятся к разным семантическим классам и имеют в корпусе разную разметку: *атака* имеет пометы 'движение', 'война'; *маскарад, митинг* – 'мероприятие', *лекция* – 'мероприятие', 'обучение', *матч* – 'мероприятие', 'спорт', *концерт* – 'мероприятие', 'искусство', *дуэль* – 'взаимодействие' и т.д. В составе оборотов эти семантические различия отчасти нейтрализуются (лишь отчасти – потому что семантика существительных, конечно, влияет на языковое поведение оборотов), и в словаре соответствующие обороты (*в атаке, на митинге, на матче* и т.д.) получают единую помету СИТУАЦИЯ (с последующей детализацией по подклассам).

Эти обороты обозначают определенные отрезки (интервалы) с определенным типом занятости – фрагменты жизни человека. Жизнь разных людей состоит из таких видов деятельности и времяпрепровождения, как работа, учеба, походы в кино, поездки в командировку, в отпуск, на рыбалку и т.п. Разумеется, у разных людей – разные наборы и пропорции разных видов деятельности, существенно, однако, что это конвенциональные типы занятости. Ситуации, которые обозначаются такими оборотами, в данном обществе являются общепринятыми (или достаточно распространенными), а их содержание – общеизвестным. Не все такие ситуации являются видами занятости в собственном смысле (*на работе, на тренировке*), это могут быть и развлечения (*на дискотеке*), а могут быть, наоборот, «принудительные» состояния (*в плену, в заключении*). Интервалы занятости могут обозначаться собственно ситуативными существительными (*на работе, на танцах, на параде, на свадьбе*), а могут – существительными-названиями учреждений, которые в составе оборота метонимически обозначают ситуацию: *в театре* (= *на спектакле*), *в школе, в институте* (= *на занятиях, на уроках*), *в тюрьме* (отбывает срок заключения), *в магазине* (делает покупки), *в ресторане* (ест), *в парикмахерской* (стрижется) и т.д. Хотя эти существительные с точки зрения корпусной разметки не относятся к классу непредметных, в словаре оборотов они (т.е. обороты с ними) попадают в один класс с непредметными существительными. Совершенно не случайно, что в русском языке так много оборо-

тов, обозначающих конвенциональные типы занятости. Далеко не для всех ситуаций есть специальные глаголы, но если и есть, они, как правило, не вмещают того объема информации, который связан с оборотом. Например, оборот «на рыбалке» нельзя приравнять к глаголу «рыбачить» или «удить рыбу», т.к. оборот вмещает в себя целый «пакет» более или менее конвенциональных действий, из-за чего и является особой словарной единицей. То же относится и к предметным существительным в ситуативном значении. Композиционное значение локативных сочетаний *в ресторане* или *в магазине* – '(X) находится в помещении ресторана / магазина' (например, если человек, спасаясь от погони, забежал в ресторан, то он 'в ресторане'). Однако оборот (*Он каждый вечер в ресторане; Где мама? – В магазине*) имеет гораздо более богатое и сложное значение (включающее, конечно, и компонент 'находиться в помещении ресторана / магазина').

И по семантическим и по синтаксическим свойствам ситуации максимально близки к полноценным глагольным предикатам. Хотя обороты, в отличие от глаголов, не имеют парадигмы словоизменения, они имеют ПАРАДИГМУ УПОТРЕБЛЕНИЙ, сопоставимую с парадигмой употреблений глагольных форм. Глагол употребляется в личной форме как сказуемое; в форме деепричастия – как обстоятельство (времени: *Уходя из дома, забыл выключить свет* ('когда уходил'); причины: *Испугавшись медведя, охотники бросились бежать* ('так как испугались'), условия: *Доверившись безответственному человеку, не жди ничего хорошего* ('если доверишься') и др.⁶ (соответствующие значения могут передаваться и обстоятельственными придаточными, но тогда уже функцию выражения этих значений берет на себя не глагольная форма, а союз); в форме причастия – как определение. Аналогично – обороты-ситуации употребляются:

– в функции сказуемого при личном существительном (разумеется, со связкой): *Вчера я был с Валей на футболе;*

– в обстоятельственной функции – в качестве так наз. детерминанта («при предложении»): *На футболе я сломал четвертый палец на левой ноге; Вчера на футболе я видел артиста*

⁶ Богуславский И.М. О семантическом описании русских деепричастий: неопределенность или многозначность // Известия АН СССР. Сер. Лит. и яз. М., 1977, т. XXXVI, № 3.

Крючкова ('когда был на футболе'); в позиции при глаголе: *Не мог же я признаться, что сломал палец на футболе;*

– в атрибутивной функции: *Человек на войне и человек в плену – это два разных человека.*

Эмоции. Другой большой класс существительных с развитой парадигмой – эмоции. Точнее, у этих существительных – две парадигмы: одна – парадигма самих конструкций, причем у разных эмоций разный набор предложных конструкций (ср.: *в ужасе, с ужасом, от ужаса, к ужасу* (кого); *в злобе, со злобой, без злобы, от злобы, из злобы, со злобы, по злобе*); другая – парадигма отдельной конструкции, т.е. возможные для данной конструкции синтаксические функции и синтаксические позиции. Основной является *V*-конструкция, т.к. только она может выступать в позиции **предиката**, ср. *Врач в негодовании; Дед был в печали*; возможны также позиции **ко-предикатива**: *Вы больше не проснетесь утром в грусти и раздражении; Петр сидел в разочаровании*; прилагательного **обстоятельства**: *Какой там кот? — в злобе закричала девица, — осёл у нас в филиале сидит!* Обороты с другими предлогами – *С, БЕЗ, ОТ, ИЗ, ПО, К* – не бывают предикатами, а бывают только обстоятельствами (или несогласованными определениями), зато они выражают значения, нехарактерные или даже невозможные для *V*-конструкций, – образ действия, причина; метатекстовое (вводная конструкция с предлогом *к*): *Я только любитель, — повторил он с грустью; Ловко вложил ему в руку целковый, принятый спокойно, без благодарности; Потом почти плакал от грусти и от бессилия; Некоторые меня отговаривали писать из злобы, другие из зависти; А если и ярился на вас, то не по злобе; И со злобы эту бутылку об стену как шарахнул; Мы, к сожалению, находимся скорее в обороне, чем в наступлении.* Некоторые конструкции могут употребляться при прилагательном (ср. *необузданные в страсти женщины; злые от грусти глаза*), а иногда и при существительном (*женщины с непередаваемой грустью в глазах*). Кроме того, конструкции могут иметь «расширенную» реализацию – с прилагательным, ср. *с глубокой благодарностью, с невыразимой грустью*; с заполненной валентностью существительного: *в злобе на студию; от беспокойства за нее.*

В заключение приведем некоторые типы запросов, с которыми пользователь может обращаться к электронному словарю:

– списки конструкций с разными предлогами, а также конструкций с беспредложными обстоятельственными и атрибутивными падежами – творительным (ср. *авансом, прыжками, угрозами*) и родительным (ср. [каких] *принципов*, [какого] *профиля*);

– семантические классы и подклассы конструкций. Поскольку в словаре обороты получают собственную разметку по классам, но при этом существительные внутри оборота сохраняют и исходную корпусную семантическую разметку, это позволит пользователю делать наблюдения и обобщения относительно языкового поведения существительных разных классов (например, в причинной конструкции с предлогом *по* участвуют не только эмоции, но и существительные других семантических классов, ср. *по ошибке, по незнанию, по легкомыслию*; а существительные со значением речи или информации с предлогом *по* образуют конструкцию с другим – непричинным – значением, ср.: *по сообщению, по информации, по отзывам, по показаниям кого*);

– парадигма существительного. На каждое существительное данного класса будет выдаваться таблица, включающая полный набор конструкций, набор синтаксических позиций (функций) для каждой конструкции и представительные примеры из Корпуса. В рамках этих парадигм могут изучаться закономерности сочетания с теми или иными предлогами, типы семантических и синтаксических изменений (утрата валентности, ср. *сомнения в чем – без сомнений согласился*), процессы грамматикализации (превращение оборота в производный предлог, ср.: *по воле кого; с согласия кого; в расчете на что; в интересах кого / чего* и т.п.) и фразеологизации (ср.: *в конце концов; в первом приближении; на свой страх и риск*).

В техническом смысле разметка оборотов не требует дополнительной разметки в НКРЯ. Разметка осуществляется в электронном словаре, каждый вход которого имеет ссылку на Корпус. Т.е. пользователь получает лингвистическую характеристику оборота из словаря и примеры из корпуса на любой оборот в режиме онлайн (опыт создания таких электронных словарей с прямым выходом в Корпус уже имеется, см. словари, созданные на основе НКРЯ, на сайте Института русского языка им. В.В. Виноградова РАН: <http://dict.ruslang.ru/>).